

LINKgroup

Distance Learning System

Predviđanje

Python Data Access

Kombinacije

- Ukoliko redosled elemenata nema značenje, radi se o **kombinacijama** elemenata
- Na primer, brojevi 1 i 2, mogu dati uvek samo jednu dvocifrenu kombinaciju jer su i 12 i 21 brojevi 1 i 2
- Od brojeva 1,2 i 3, mogu se formirati tri dvocifrene kombinacije: (1,2),(1,3) i (2,3)

$$C_{(n,r)} = \frac{n!}{r!(n-r)!}$$

Ukupno cifara Jedan set

Permutacije

- Kod permutacija položaj elemenata ima značenje
- Na primer, brojevi 1 i 2, mogu dati dva različita dvocifrena broja, 12 i 21
- Od brojeva 1,2 i 3 možemo napraviti 6 dvocifrenih kombinacija: (1,2), (1,3),(2,1),(2,3),(3,1),(3,2)

$$P_{(n,r)} = \frac{n!}{(n-r)!}$$

Ukupno cifara Jedan set

Ponavljjanje

- Kombinacije i permutacije se mogu pojavljivati sa ili bez ponavljanja vrednosti
- Ukoliko postoji ponavljanje, broj opcija se dodatno povećava
- Na primer, za brojeve 1,2 i 3, možemo dobiti sledeće dvocifrene kombinacije:
(1,1),(1,2),(1,3),(2,1),(2,2),(2,3),(3,1),(3,2),(3,3)

$$P_{(n,r)} = n^r$$

Ukupno cifara Jedan set

Verovatnoća

- Verovatnoća je šansa da će se nešto dogoditi, najčešće izražena u brojevima
 - Verovatnoća se predstavlja postotkom, odnosno, brojem između nula i jedan
 - U najosnovnijem obliku, verovatnoća da će se nešto dogoditi jeste jedan (ili broj pozitivnih opcija) podeljeno sa ukupnim brojem opcija
-
- Šansa da će prilikom bacanja novčića pasti pismo jeste $1/2$
 - Šansa da prilikom bacanja kockice dobijete određeni broj, jeste $1/6$

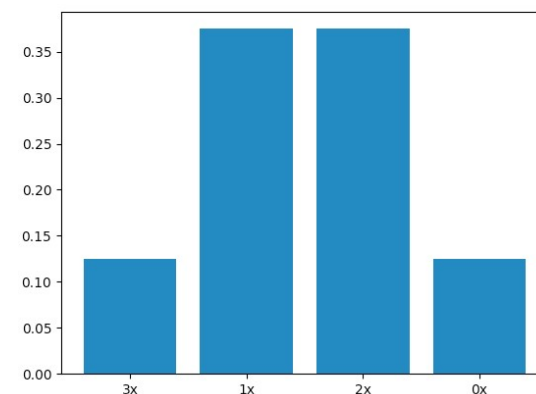


Izračunavanje verovatnoće

- Iako znamo da je 0.5 šansa da će na novčiću pasti glava, najčešće nas interesuju šanse za određenu sekvencu događaja. Na primer, da dva puta padne glava, jednom pismo i slično
- Ukoliko bacimo novčić tri puta:

- Broj kombinacija
 - Da tri puta bude pismo, šanse su:
 - $1 / 8 = 0.125$
 - Da jednom bude pismo, šanse su:
 - $3 / 8 = 0.375$
 - Da dva puta bude pismo, šanse su:
 - $3 / 8 = 0.375$
 - Da ni jednom ne bude pismo, šanse su:
 - $1 / 8 = 0.125$

$$2^3 = 8$$



Predviđanje

- Na osnovu verovatnoće, možemo predvideti sledeći korak u nekoj sekvenci koraka
- Na primer, ako je novčić bačen dva od tri puta i dva puta je dobijeno pismo, imamo zapravo 50% šansi da dobijemo tri pisma
- Ako runda tek počinje, sistem zna da su šanse za dobitak tri pisma tek 0.125

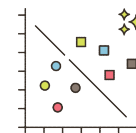
Predviđanje iskustvom

- Za situacije u kojima postoje matematički predvidivi ishodi moguće je koristiti i matematičku verovatnoću
- Za situacije u kojima ne postoji matematička logika kombinacijama eksperimentalnih vrednosti i ishoda, možemo formirati linearnost na osnovu iskustva
- Predviđanja obavljena na ovaj način, spadaju u kategoriju **mašinskog učenja**

Varijante mašinskog učenja

- Nadzirano

Regresija
Klasifikacija



- Nenadzirano

Klasterovanje



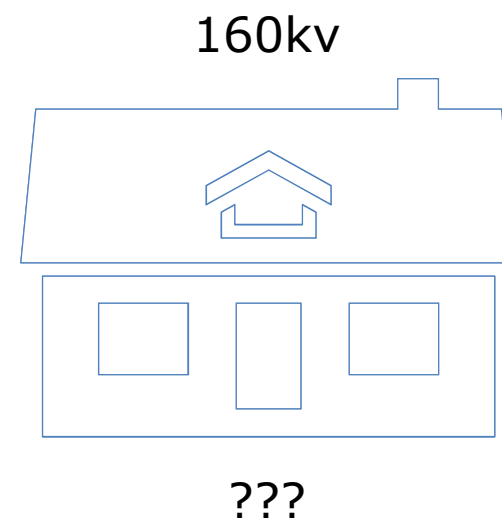
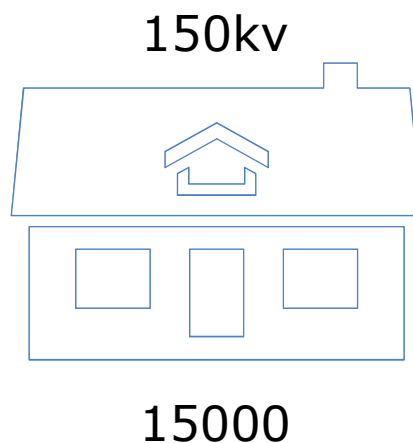
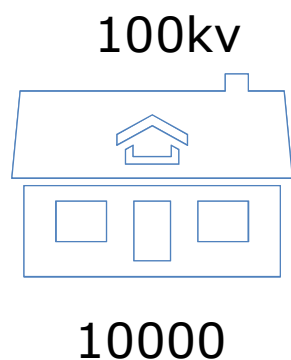
- Ojačano

Neuronska mreža



Regresija

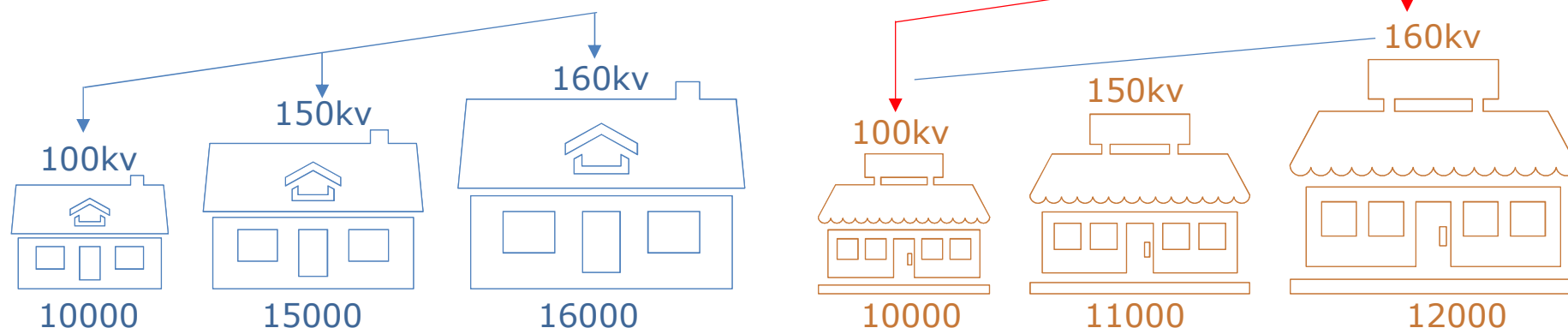
- Na osnovu iskustava iz "prošlosti", predviđamo "budućnost"



- Predviđanje može biti **kontinualno** i **diskretno** (kategoričko)

Bias i varijansa

- **Bias** - Odstupanje od predviđenih i stvarnih podataka
 - Underfit - Model nije dobar
- **Variance** - Razlika u rezultatima predikcija nad različitim setovima podataka
 - Overfit - Model je previše dobar za jedan set podataka, ali ne i za ostale



Unakrsna validacija

- Postoji mnogo algoritama mašinskog učenja i ne odgovara svaki od njih svakoj kolekciji podataka, čak i ukoliko je i sama kategorija algoritma odgovarajuća
- Kvalitet algoritma se može proveriti tehnikom unakrsne validacije
- Kod unakrsne validacije, podatke delimo na dva dela, od kojih je jedan veći i koristi se za treniranje, dok je drugi manji i koristi se za proveru dobijenih rezultata

Uspešnost klasifikacije

- Matrica zabune
- Tačnost
 - Koliko je često tačno predviđanje
- Preciznost
 - Koliko je tačno predviđanje za klasu
- Odziv
 - Koliko je puta on što je stvarno tačno bilo predviđeno tačno
- F rezultat
 - Balansirani prosek između odziva i preciznosti
- Površina ispod ROC (Receiver Operating Characteristics) krive

Matrica zabune

		Predvidjeno	
		Da	Ne
Stvarno	Da	TP	FN
	Ne	FP	TN

TP = True positive
FP = False positive
TN = True Negative
FN = False negative

Ako smo za **15** slučajeva predvideli
8 slučajeva **da**, a **7** slučajeva **ne** i
pri tom, pogodili **3** slučaja za **da** i **4**
slučaja za **ne**

n=15	Da	Ne
Da	3	3
Ne	5	4

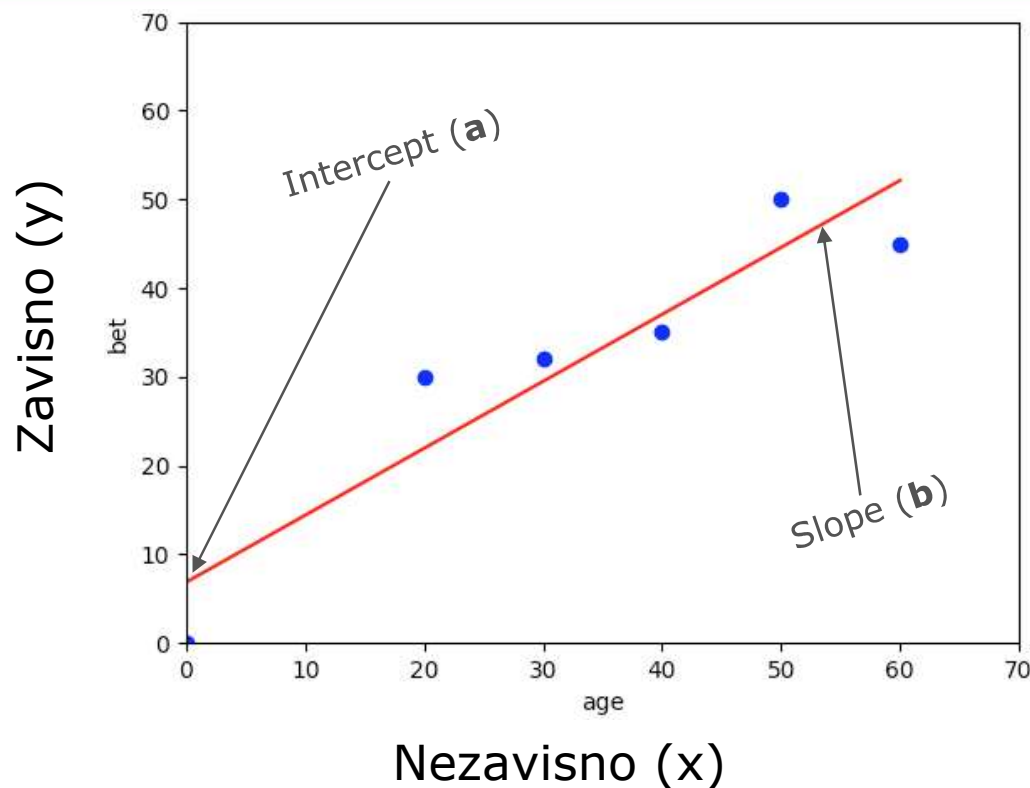
Upotreba matrice zabune

n=15	Da	Ne	
Da	3	3	6
Ne	5	4	9
	8	7	

- Tačnost: $TP+TN / n$ $(3+4) / 15 = 0.47$
- Preciznost (za P): $TP / (TP+FP)$ $3 / 8 = 0.38$
- Preciznost (za N): $TN / (TN+FN)$ $4 / 7 = 0.57$
- Odziv (za P): $TP / (TP+FN)$ $3 / (3+3) = 0.5$
- Odziv (za N): $TN / (TN+FP)$ $4 / (4+5) = 0.44$
- F rezultat (za P):
 $(2 * ((odziv * preciznost) / (odziv + preciznost)))$ $2 * ((0.5 * 0.38) / (0.5 + 0.38)) = 0.43$

Linearna regresija

(pdap-ex01 linregression.py)



Na osnovu prethodnih ulaza i izlaza predviđamo buduće izlaze

Na primer, ako tridesetogodišnjaci ulažu po 30, četrdesetogodišnjaci po 40, a pedesetogodišnjaci po 50, korisnik od 60 godina će verovatno uložiti 60

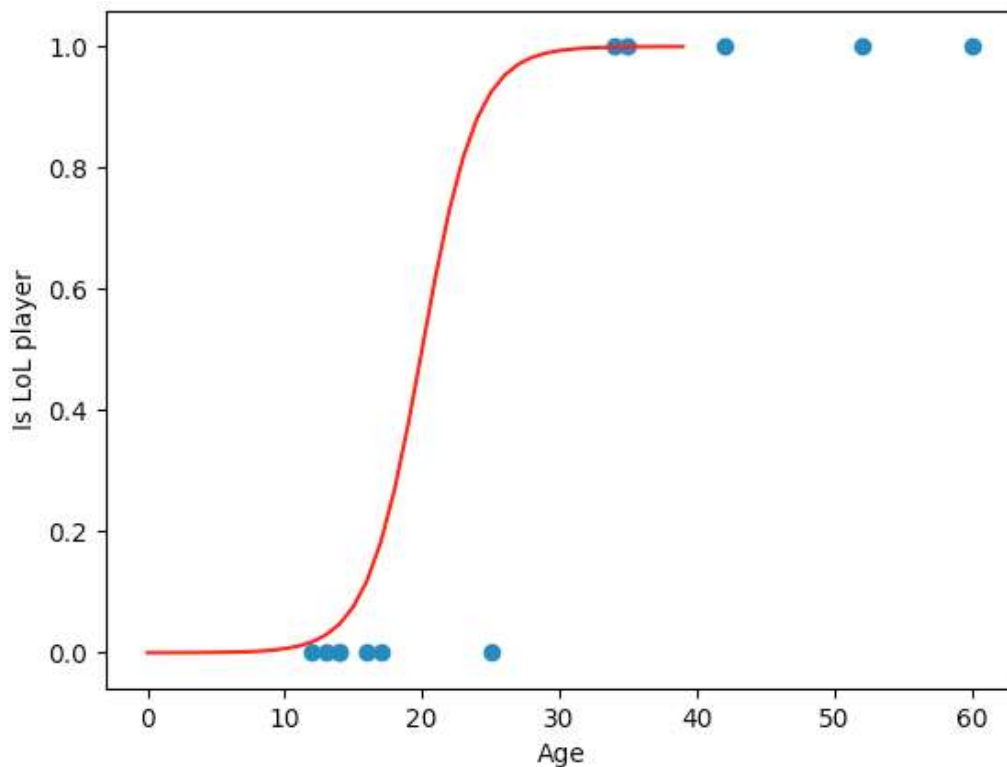
$$a = \frac{(\sum y \sum x^2) - (\sum x \sum xy)}{n(\sum x^2) - (\sum x)^2}$$

$$b = \frac{n(\sum xy) - (\sum x \sum y)}{n(\sum x^2) - (\sum x)^2}$$

$$y = a + (b * x)$$

Logistička regresija

(pdap-ex01 logregression.py)



- Kod ovog modela izlazna vrednost je binarna dok su ulazne vrednosti kontinualne
- Sigmoid funkcija se primenjuje na a i b dobijene u formuli za linearnu regresiju

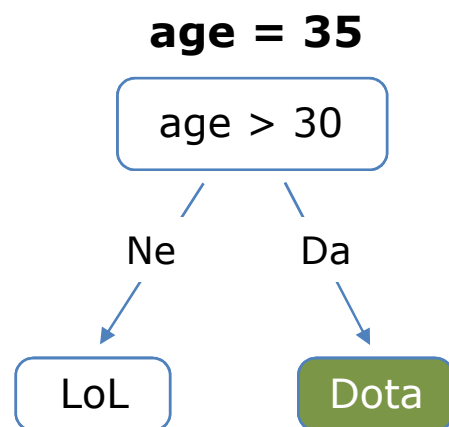
$$y = a + (b * x)$$

$$\text{sig} = 1 / (1 + \exp(-y))$$

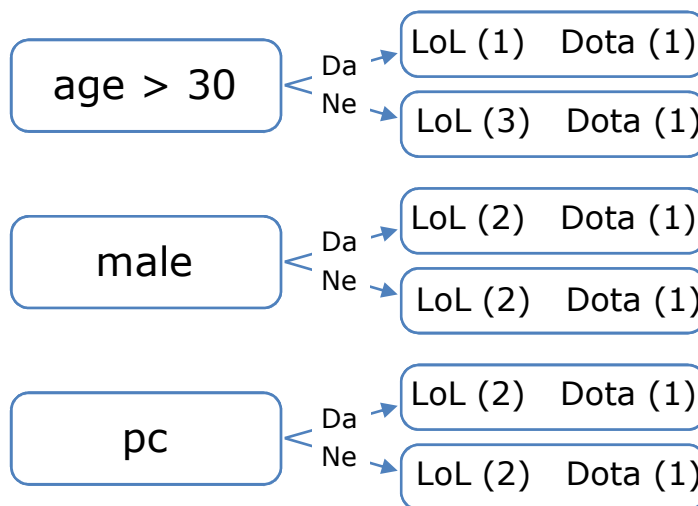
$$\mathbf{y} = \log(\text{sig} / (1 - \text{sig}))$$

Decision tree

age	gender	platform	game
35	m	macos	Dota
25	m	macos	LoL
18	f	pc	Dota
23	f	pc	LoL
27	m	macos	LoL
32	f	pc	LoL



Da li je lol igrač?



$$\begin{aligned}
 &1 - (1 / (1+1))^2 - (1 / (1+1))^2 = 0.5 \\
 &1 - (3 / (3+1))^2 - (1 / (3+1))^2 = 0.375 \\
 &= (2 / (2+4)) * 0.5 + (4 / (2+4)) * 0.375 \\
 &= 0.16 + 0.25 = \mathbf{0.41}
 \end{aligned}$$

$$\begin{aligned}
 &1 - (2 / (2+1))^2 - (1 / (2+1))^2 = 0.667 \\
 &1 - (2 / (2+1))^2 - (1 / (2+1))^2 = 0.667 \\
 &= (3 / (3+3)) * 0.667 + (3 / (3+3)) * 0.667 \\
 &= 0.3335 + 0.3335 = \mathbf{0.667}
 \end{aligned}$$

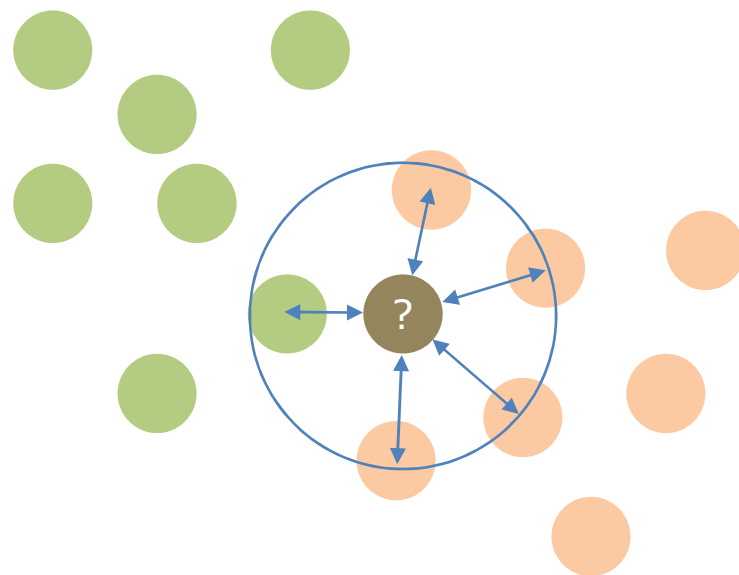
$$\begin{aligned}
 &1 - (2 / (2+1))^2 - (1 / (2+1))^2 = 0.667 \\
 &1 - (2 / (2+1))^2 - (1 / (2+1))^2 = 0.445 \\
 &= (3 / (3+3)) * 0.667 + (3 / (3+3)) * 0.667 \\
 &= 0.3335 + 0.3335 = \mathbf{0.667}
 \end{aligned}$$

Decision tree klasifikacija

LINKgroup

K-Nearest Neighbours

(pdap-ex01 knn)



K-nearest neighbour Klasifikacija

LINKgroup

Naive Bayes

(pdap-ex01 naivebayes)

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)}$$

Practice over 3h per day	Play tank	Play with premade	Win
1	1	1	1
1	0	1	1
0	0	1	0
1	0	0	0
1	1	0	1
1	0	1	1
1	1	0	1
1	1	1	1
0	0	0	0
0	1	0	0
1	0	0	0

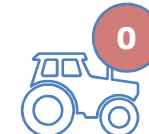
Pobede

Porazi

6 / 11

5 / 11

Ulaz:



1 : 6 / 6 | 2 / 5

0 : 0 / 6 | 3 / 5



1 : 4 / 6 | 1 / 5

0 : 2 / 6 | 4 / 5



1 : 4 / 6 | 1 / 5

0 : 2 / 6 | 4 / 5

Pobeda: $(6 / 6) * (2 / 6) * (4 / 6) * (6 / 11) = 0.121$

Poraz: $(2 / 5) * (4 / 5) * (1 / 5) * (5 / 11) = 0.0288$

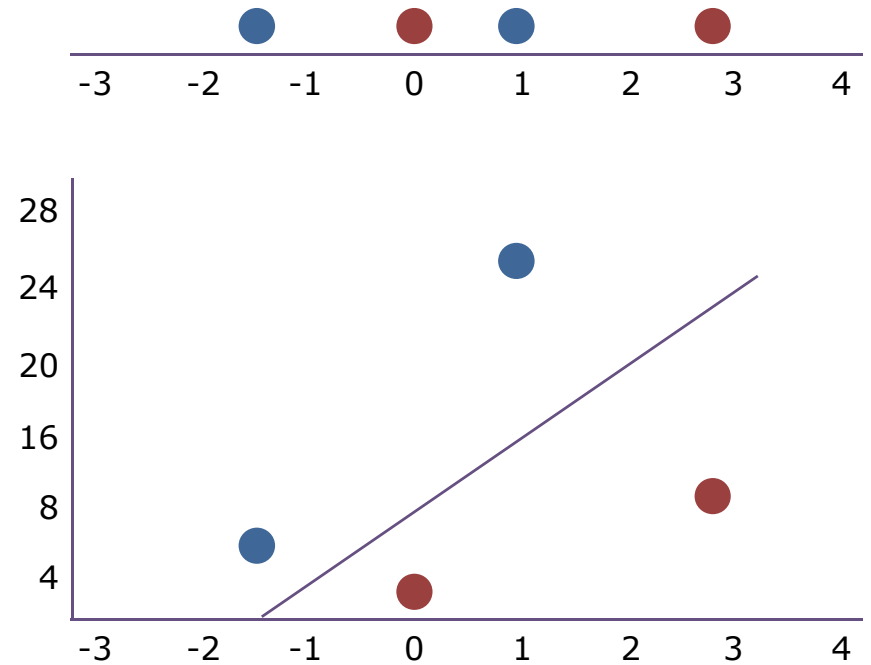
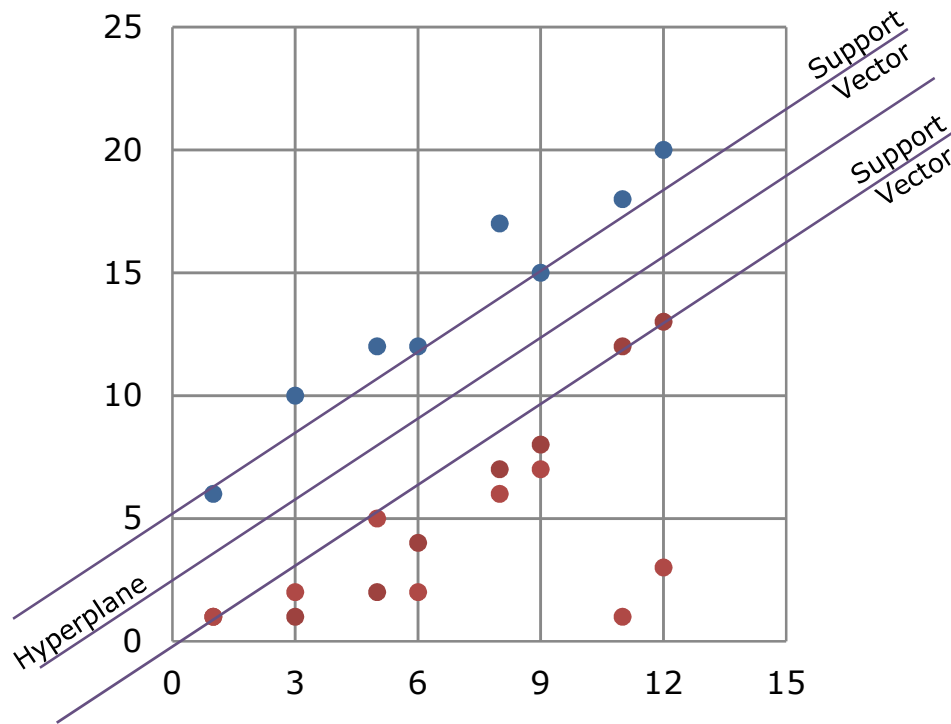
px: $(8 / 11) * (6 / 11) * (5 / 11) = 0.180$

Pobeda: $(0.121 / 0.180) > (0.0288 / 0.180)$
0.67
0.16

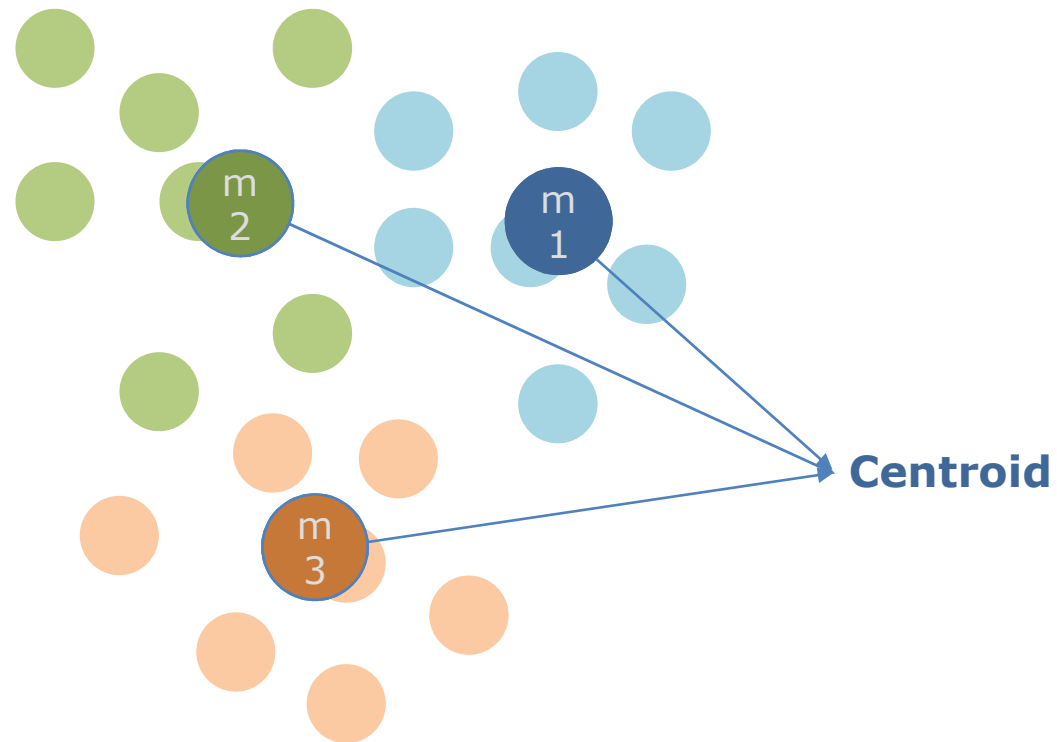
Klasifikacija

LINKgroup

Support vector machine

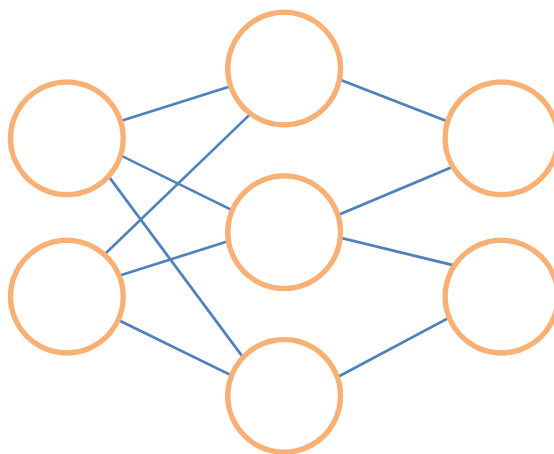


K-means



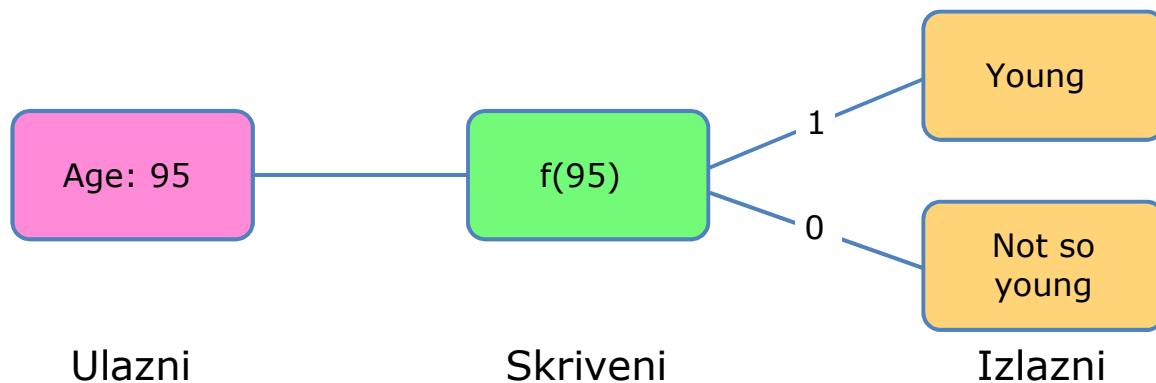
Veštačka neuronska mreža

- Veštačka neuronska mreža je skup uzajamno povezanih čvorova koji imaju različite, najčešće numerički predstavljene, karakteristike
- Veštačka neuronska mreža se može realizovati u različitim oblicima (Feed Forward, Radial, Reccurent, Convolutional, Modular...)



Struktura veštačke mreže

- Najmanji sastojak neuronske mreže jeste jedan čvor (neuron)
- Čvor može imati različitu strukturu u zavisnosti od svoje pozicije u mreži
- Pozicije u mreži mogu biti: **ulazne**, **skrивene** i **izlazne**
- Mreža ima jedan sloj ulaznih čvorova i jedan sloj izlaznih čvorova, dok može biti više slojeva sa skrivenim čvorovima



Credits



<https://www.flaticon.com/authors/freepik>



<https://www.flaticon.com/authors/flat-icons>



<https://www.flaticon.com/authors/nikita-golubev>



<https://www.flaticon.com/authors/becris>



<https://www.flaticon.com/authors/catkuro>

LINKgroup